

01 Jan 1999

# A Biologically Inspired Connectionist Model for Image Feature Extraction in 2D Pattern Recognition

Cihan H. Dagli
*Missouri University of Science and Technology*, dagli@mst.edu

Raymond K. Chafin



## Recommended Citation

C. H. Dagli and R. K. Chafin, "A Biologically Inspired Connectionist Model for Image Feature Extraction in 2D Pattern Recognition," *Proceedings of the International Joint Conference on Neural Networks, 1999*, Institute of Electrical and Electronics Engineers (IEEE), Jan 1999.
The definitive version is available at https://doi.org/10.1109/IJCNN.1999.833506

# A Biologically Inspired Connectionist Model for Image Feature Extraction in 2D Pattern Recognition

Raymond K. Chafin, Smart Engineering Systems Laboratory, University of Missouri – Rolla, Rolla, MO 65409-1060, chafin@advertisnet.com
Cihan H. Dagli, Smart Engineering Systems Laboratory, University of Missouri – Rolla, Rolla, MO 65409-1060, dagli@umr.edu

## Abstract

*A new edge detection method is presented which borrows from recent research into primate vision biology, and offers improved noise performance over classical methods. Beginning with spatio-temporal shunting models for retinal cones, horizontal cells, bipolar cells, and retinal ganglions, a set of simplified steady-state solutions are developed which lend themselves to efficient computation on standard computer equipment. The retinal model output is found to be nominally equivalent to the classical edge detector, but is produced differently. Developed somewhat speculatively from incomplete biological information, a simplified model of the Lateral Geniculate Nucleus (LGN) has been produced. Taking the output of the retinal model, the LGN simple cell and interneuron models perform noise reduction and segment completion. An orienting subsystem is used to adaptively infer segment strengths and orientations, throwing out spurious and foreshortened edges, while retaining and filling in the longer edges.*

## Introduction

One pressing problem in pattern recognition applications is the preliminary image processing to extract the relevant features of images containing candidate objects for recognition. The performance of the edge detector can in large part determine the overall performance of a pattern recognition system. This is particularly true when the objects to be recognized are three dimensional soft shapes, whose edges are used to map the 3D contours of the candidate objects into 2D space. The results can be disappointing, with broken contours and much extraneous clutter due to lighting irregularities, camera noise, and surface blemishes in the objects.

Rudimentary analysis of the edge detection process reveals that all edge detectors amount to various forms of high-pass filters. In the presence of broadband or white noise at the input, these filters have the characteristic of amplifying the noise variance. The traditional approach to overcoming the noise problem has been to tailor the detection filter response. More recently, biologically-inspired edge detection models have shown promise for noise reduction at the expense of susceptibility to optical illusions.

## Edge Detection and Noise

When three dimensional objects are presented to a camera, the projection to the camera's focal plane is a two dimensional intensity map that combines the factors of viewing angle, lighting angle, object topography, and object surface texture. One view of this result is to consider intensity at any point on the focal plane to be a product of the luminance and the reflectance. The largest changes in luminance will occur when object topography causes shadowing of the light source or there are large changes in the angular relationships between light source and the object surface. Reflectance changes occur most strongly with region to region changes in the surface textures of the objects being viewed. Mathematically, this is expressed in equation (1) below.

$$f(x,y) = l(x,y)\rho(x,y) \quad (1)$$

where: $f(x,y)$ is the received light intensity, $l(x,y)$ is the luminance, and $\rho(x,y)$ is the reflectance

In the types of pattern recognition systems under consideration, the goal is to determine the boundaries of the objects and of the features within the objects themselves. The shapes of these boundaries are sufficiently unique that one may recognize an object with a high degree of confidence from this information alone. The goal of edge detection is to isolate and localize these boundaries, while eliminating extraneous clutter and noise. To do this, we will consider the two principal forms of

0-7803-5529-6/99/$10.00 ©1999 IEEE

classical edge detectors, gradient-based and Laplacian-based. Gradient-based detectors operate by producing output responses proportional to the rate of change in the intensity image, and are capable of returning edge orientations. Equations (2) through (6) express this mathematically. Edge locations are ordinarily determined by peak-finding or by thresholding and thinning.

$$\nabla f = \frac{\partial f}{\partial x} x + \frac{\partial f}{\partial y} \bar{y} \tag{2}$$

$$\frac{\partial f}{\partial x} = l(x,y)\frac{\partial \rho}{\partial x} + \rho(x,y)\frac{\partial l}{\partial x} \tag{3}$$

$$\frac{\partial f}{\partial y} = l(x,y)\frac{\partial \rho}{\partial y} + \rho(x,y)\frac{\partial l}{\partial y} \tag{4}$$

$$|\nabla f| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2} \tag{5}$$

$$\theta_f = \tan^{-1}\left(\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial y}\right) \tag{6}$$

where: $|\nabla f|$ is the edge gradient and $\theta_f$ is the edge orientation

Laplacian-based edge detectors operate by taking the Laplacian of the intensity image, as expressed in equations (7) through (9). Edge locations are determined by the zero-crossings of the output, and edge orientations are not available.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \tag{7}$$

$$\frac{\partial^2 f}{\partial x^2} = l(x,y)\frac{\partial^2 \rho}{\partial x^2} + 2\frac{\partial l}{\partial x}\frac{\partial \rho}{\partial x}\rho(x,y)\frac{\partial^2 l}{\partial x^2} \tag{8}$$

$$\frac{\partial^2 f}{\partial y^2} = l(x,y)\frac{\partial^2 \rho}{\partial y^2} + 2\frac{\partial l}{\partial y}\frac{\partial \rho}{\partial y}\rho(x,y)\frac{\partial^2 l}{\partial y^2} \tag{9}$$

We observe that the differentiations used in detecting edges are equivalent to applying a high-pass filter with a cutoff at DC. If the scene contains additive uncorrelated noise that has been bandlimited by the imaging system or by explicit filtering, we can gain some insight into the noise performance of the two edge detection methods through a one dimensional analysis. We assume the original scene contains noise which is filtered at cutoff frequency $\omega_0$ using an ideal low-pass filter. The variance of the noise in the image is assumed to be $\sigma_n^2$. The power spectral density of the image noise is given by:

$$S_{nn}(\omega) = \begin{cases} \dfrac{\sigma_n^2}{2\omega_0} & \text{for } |\omega| \le \omega_0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

The image gradient and Laplacian power spectral densities are given by equations (11) and (12).

$$S_{gg} = \begin{cases} \omega^2 \dfrac{\sigma_n^2}{2\omega_0} & \text{for } |\omega| \le \omega_0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

$$S_{ll} = \begin{cases} \omega^4 \dfrac{\sigma_n^2}{2\omega_0} & \text{for } |\omega| \le \omega_0 \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

The gradient and Laplacian edge image noise variance are computed by:

$$\sigma_g^2 = \int_{-\infty}^{\infty} S_{gg}(\omega)\, d\omega \tag{13}$$

$$\sigma_l^2 = \int_{-\infty}^{\infty} S_{ll}(\omega)\, d\omega \tag{14}$$

$$\sigma_g^2 = \frac{\omega_0^2}{3}\sigma_n^2 \tag{15}$$

$$\sigma_l^2 = \frac{\omega_0^4}{5}\sigma_n^2 \tag{16}$$

The noise variance at the output of the gradient edge detector is amplified by the square of the filter bandwidth. The Laplacian edge detector noise variance is amplified by the fourth power of the filter bandwidth. Traditional edge detection methods have sought optimal filters to reduce the noise. However, we shall see that biological vision systems employ more sophisticated methods.

## Retinal Models

The retina is a complex structure comprised of six major cell types, two of which are photoreceptors. Rod cells are sensitive to broadband light, with a peak at 498 nm. They are distributed throughout the retina, with a thinner concentration at the fovea, and are connected with a high degree of parallelism. Cone cells are responsible for color vision, and come in three types, each employing a different photochemical. These photochemical differences are responsile for peak sensitivities to light at 420 nm, 534 nm, and 564 nm [7]. Cone cells are concentrated most densely within the fovea and unlike the rods, are not connected with the same high degree of parallelism.

Rod and cone photoreceptors are synapsed with bipolar cells and horizontal cells, which are responsible for the contrast enhancement characteristics of the retina resulting from the lateral inhibition characteristic of the photoreceptor connections.

Amacrine cells play a role in modulating the outputs of the bipolar and horizontal cells. Ganglion cells connect the retinal output to the optic nerve. P type ganglion cells

synapse to the parvocellur LGN structure, while M type ganglion cells synapse to magnocellular LGN structures. Unlike the first four cell type, amacrine and ganglion cells fire action potentials. When an input stimulus is reaches a threshold, the cell fires a burst of energy, resets, and fires again when the threshold is again reached. Rods, cones, bipolar cells, and horizontal cells, all transmit a graded nonlinear response.

Lateral inhibition in the retinal bipolar and horizontal cells is described by the basic push-pull shunting equation (17). $W^+$ and $W^-$ are the excitation and inhibition weight matrices which are multiplied by the input image neighborhood to yield the network gross excitation and inhibition. Adjustment of the weight matrices can yield either an on-center/off-surround characteristic or an off-center/on-surround characteristic. The network gradually moves toward a steady state value at a rate determined by the time constant $\varepsilon_1$, and the steady state value can be readily determined by setting the right hand side of the equation to zero.

$$\varepsilon_1 \frac{\partial n_1}{\partial t} = -n_1 + (b_1^+ - n_1) W_1^+ p_1 - (n_1 + b_1^-) W_1^- p_1 \quad (17)$$

The retina layer 1 model begins with excitations coming from the cone cells, and includes the bipolar and horizontal cells. Rod cells are ignored as they will be in the layer 2 model which follows. At the layer 1 stage, the contributions of the amacrine cells and the ganglions are ignored. The shunting model equations governing the bipolar and horizontal cells are given by equation (18). The steady state solution is given by equation (19).

$$\frac{dx(t)}{dt} = -Px(t) + [Q - x(t)]e(t) - [R + x(t)]i(t) \quad (18)$$

$$x = \frac{Qe - Ri}{e + i + P} \quad (19)$$

In equations (18) and (19), $x(t)$ represents the activation of the neuron, $e(t)$ the total excitatory input, and $i(t)$ the total inhibitory input. The constants $P$, $Q$, and $R$ represent the rate of passive decay, excitatory saturation point, and inhibitory saturation point, respectively. Since the decay parameter is unnecessary in a steady state solution, it is retained and set to a small value to prevent divide by zero problems. By defining a neighborhood span $n_k$, and an activation threshold $\theta_b$, the total bipolar cell activation is computed [2].

$$b(x,y) = \frac{Q_b c(x,y) - R_b \sum_{i=x-n_k}^{x+n_k} \sum_{j=y-n_k}^{y+n_k} c(i,j)}{c(x,y) + \sum_{i=x-n_k}^{x+n_k} \sum_{j=y-n_k}^{y+n_k} c(i,j) + P_b} \quad (20)$$

$$b(x,y) = 0 \text{ if } b(x,y) \le \theta_b \quad (21)$$

$$R_b = \frac{1}{(2n_k + 1)^2} \quad (22)$$

Parameters $P_b$, $Q_b$, and represent the bipolar cell rate of passive decay, excitatory saturation point, and inhibitory saturation point, respectively. Quantity $c(x,y)$ is the cone cell activation at location (x,y).

Simulations were performed using a neighborhood span of 1 with $Q_b$ set to 1, $P_b$ set to 0.01, and $\theta_b$ set to 0.02. The input image was a face, since it had wide range of spatial frequencies and noise. The input image appears in figure 2. The negative image of the bipolar cell output appears in figure 3. As can be seen, the contrast has been enhanced to the point of the retina behaving as an edge extractor.

Figure 1. Input image

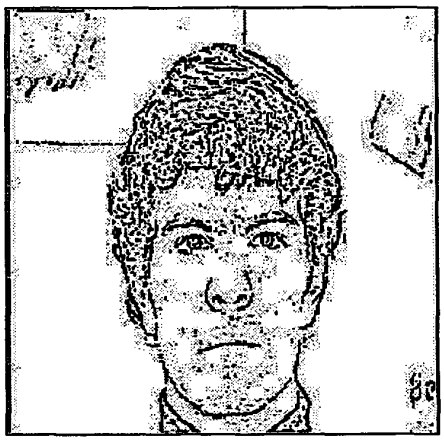

Figure 2. Retina layer 1 output

We now extend the retina layer 1 model to include the effects of cone-to-cone gap junction thresholding. For bipolar cells whose excitations exceed a threshold, their

central cone excitations leak out to the surrounding cones. This imparts a blur to the image. The blur computation is given by equations (23) and (24).

$$c_{t+1}(x,y) = c_t(x,y) + \sum_{i=x-1}^{x+1} \sum_{j=y-1}^{y+1} M[c_t(i,j) - c_t(x,y)] \quad (23)$$

$$M = \begin{cases} \frac{1}{A} & \text{if } b(x,y) > \theta_f \\ 0 & \text{otherwise} \end{cases} \quad (24)$$

$c_{t+1}(x,y)$ is the photoreceptor response after blurring at location (x,y). $A$ is a constant used to determine the contribution of each gap junction to the new cone activation. $\theta_f$ is the gap threshold dependent on the bipolar cell activation.

We also incorporate into the model two types of amacrine cells. The large-field amacrine cell is governed by equations (25) through (28). Like the bipolar cell equations, the large field amacrine cell incorporates a center-surround lateral inhibition. However, since the amacrine cell fires an action potential, the model uses a log-sigmoid function to approximate a firing rate in a steady state response.

$$a_l(x,y) = R_{al} \sum_{i=x-n_{al}}^{x+n_{al}} \sum_{j=y-n_{al}}^{y+n_{al}} b(i,j) \quad (25)$$

$$R_{al} = \frac{1}{(2n_{al}+1)^2} \quad (26)$$

$$a_{l(x,y)}(i,j) = \frac{1}{1+e^{-S_{al}(net_{al(x,y)}(i,j)-T_{al})}} \quad (27)$$

$$net_{al(x,y)}(i,j) = \frac{a_l(x,y)}{b(i,j)} \quad (28)$$

$a_l(x,y)$ represents the input summation of the large field amacrine cell. $R_{al}$, $S_{al}$, and $T_{al}$ are its range, slope, and threshold parameters. $net_{al(x,y)}(i,j)$ is the feedback activation from the neighborhood at $(i,j)$ for the amacrine cell at $(x,y)$.

The small field amacrine cell is modeled in similar fashion, except for the use of a smaller neighborhood. In the interest of brevity, the equations have been omitted. For the final element of the model, the ganglion cells are described by equations (29) through (31). As before, the processing is on a neighborhood, designated here as $n_g$. A log-sigmoid function is used to approximate the firing rate of the ganglion action potential.

$$net_g = \frac{Q_g b(x,y) - R_g \sum_{i=x-n_g}^{x+n_g} \sum_{j=y-n_g}^{y+n_g} net_{as}(i,j)}{b(x,y) + \sum_{i=x-n_g}^{x+n_g} \sum_{j=y-n_g}^{y+n_g} net_{as}(i,j) + P_g} \quad (29)$$

$$R_g = \frac{1}{(2n_g+1)^2} \quad (30)$$

$$g(x,y) = \frac{1}{1+e^{-S_g(net_g(x,y)-T_g)}} \quad (31)$$

A simulation of the same image as with the layer 1 model was performed. The results (as a negative image) are shown in figure 4. The background has shifted toward the gray, but the overall noise level has been reduced without significant loss of detail.

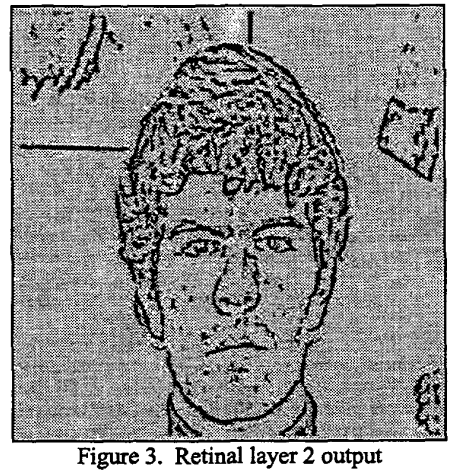

Figure 3. Retinal layer 2 output

Analysis of the model reveals that the retina layer 1 model approximates a Laplacian edge detector with thresholding added. When the effects of the amacrine and ganglion cells are included, the threshold-sensitive blurring acts as an adaptive low-pass filter. In areas where the layer 1 model exhibits a high density of edges (high noise levels), the effect is to reduce the edge densities. In areas of few edges, the detector sensitivity remains high. Computational times for the layer 2 model are prohibitively long for practical use in machine vision applications. However the results suggest that more efficient implementations may be possible.

## LGN Models

The Lateral Geniculate Nucleus consists of the parvocellular system and the magnocellular system. The parvocellular system modeled here is responsible for processing of color, texture, and shape. Of specific interest is the manner in which the LGN interacts with the visual cortex simple cells in processing object orientation Like the retinal bipolar and horizontal cells, the cells in the LGN use a lateral inhibition mechanism to produce enhanced representations of objects within the visual field. The field of retinal ganglion cells is first gated to select for a characteristic orientation before applying lateral

inhibition. Figure 4 illustrates the interconnections between the cortical simple cells, the retinal P ganglions, and the LGN interneurons.

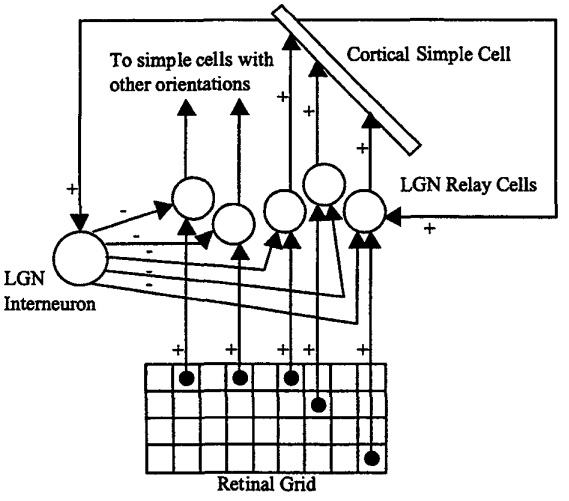


Figure 4. LGN Feedback Model

Initially, each LGN interneuron receives the direct output of the P ganglion as shown in equation (32). In the model considered here, there are four receptive fields determined from a set of 3x3 neighborhoods, 0°, 45°, 90°, and 135°. For each orientation, a recptive field is defined as shown in equation (33) [2].

$$l_{(x,y)}(t) = r_{(x,y)} \tag{32}$$

$$RF_{(x,y)k}(t) = \sum_{i=x-1}^{x+1}\sum_{j=y-1}^{y+1}\left[\mu_c T_c(i,j) - \mu_s T_s(i,j)\right] l_{(i,j)}(t) \tag{33}$$

$$T_c(i,j) = \begin{cases} 1 \text{ if } wk(i,j) = 1 \\ 0 \quad \text{otherwise} \end{cases} \tag{34}$$

$$T_s(i,j) = \begin{cases} 1 \text{ if } wk(i,j) = 0 \\ 0 \quad \text{otherwise} \end{cases} \tag{35}$$

The orientations are designated with $k=1,2,3,4$, and the parameters $\mu_c$ and $\mu_s$ are used to adjust the center and surround responses of the simple cells defined by $T_c(i,j)$ and $T_s(i,j)$. The quantity $wk(i,j)$ is the individual connection strength ay location $(i,j)$ within the local neighborhood of the $k^{th}$ orientation simple cell. Using both long-range excitatory and short-range inhibitory connections, the total activation of each simple cell at location $(x,y)$ and orientation $k$ is computed by equation (36).

$$net_{s(x,y)k}(t) = \rho_r RF_{(x,y)k}(t) + \rho_l Sl_{(x,y)k}(t) - \rho_s Ss_{(x,y)k}(t) \tag{36}$$

The terms $Sl_{(x,y)k}(t)$ and $Ss_{(x,y)k}(t)$ represent the contributions from the long-range excitatory and short-range inhibitory connections. These are calculated as follows:

$$Sl_{(x,y)k}(t) = \sum_{\substack{i=x-1 \\ i\neq j}}^{x+1}\sum_{j=y-1}^{y+1} S_{(i,j)k}(t-1) \tag{37}$$

$$Ss_{(x,y)k}(t) = \sum_{\substack{m=1 \\ m\neq k}}^{4} S_{(x,y)m}(t-1) \tag{38}$$

The parameters $\rho_r$, $\rho_l$, and $\rho_s$, are used to adjust the influence of the receptive field, excitatory long-range, and inhibitory short-range connections to the simple cell. Once the net activation for the simple cell has been calculated in equation (36), the final cell response is computed in equation (39).

$$\begin{aligned} S_{(x,y)k}(t) &= \delta_s(1-\alpha)S_{(x,y)k}(t-1) \\ &\quad + (1-\delta_s)\,\text{satlin}\left[net_{s(x,y)k}(t)\right] \end{aligned} \tag{39}$$

$$\delta_s = \begin{cases} 1 \text{ if } net_{s(x,y)k}(t) < \theta_s \\ 0 \text{ otherwise} \end{cases} \tag{40}$$

$$\text{satlin}(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 & \text{if } x > 1 \\ x & \text{otherwise} \end{cases} \tag{41}$$

If the net activation of the simple cell is below a threshold $\theta_s$, the response of the cell will decay as parameter $\alpha$. Once the simple cell responses have been determined by equation (39), the new activations for the LGN interneurons are calculated from equation (42).

$$net_{l(x,y)}(t) = r_{(x,y)} + f_p M_{cc(x,y)}(t) - f_n S_{cc(x,y)}(t) \tag{42}$$

The value $r_{(x,y)}$ is the original retinal P ganglion response at location $(x,y)$. $M_{cc(x,y)}$ and $S_{cc(x,y)}$ are the excitatory and inhibitory contributions to the individual LGN cells.

$$M_{cc(x,y)}(t) = \text{MAX}\left[S_{(x,y)k}(t)\right] \text{ for } k = 1,2,3,4 \tag{43}$$

$$S_{cc(x,y)}(t) = \sum_{i=x-1}^{x+1}\sum_{j=y-1}^{y+1} Lin_{(i,j)}(t) \tag{44}$$

$$Lin_{(x,y)}(t) = \sum_{m=1}^{4} S_{(i,j)m}(t) \tag{45}$$

Equation (43) shows the positive feedback to a LGN cell at location $(x,y)$ comes from the maximum response of the four oriented simple cells at that that location. Eventually a single orientation will dominate the feedback after some number of recursive calculations. The weighting parameters $f_p$ and $f_n$ allow for the adjustment of the positive and negative feedback. The final response of the LGN interneurons is given by equation (46).

$$l_{(x,y)}(t) = \begin{cases} \text{satlin}\left[net_{l(x,y)}(t)\right] \text{ if } net_{l(x,y)}(t) < \theta_l \\ 0 \qquad\qquad \text{otherwise} \end{cases} \tag{46}$$

The parameter $\theta_l$ is a threshold parameter for the LGN interneurons.

Using the LGN model of equations (32) through (46), a number of simulations were performed. The parameters were set as follows:

$$\mu_c = 0.3, \mu_s = 0.12$$
$$\rho_r = 0.8, \rho_l = 0.1, \rho_s = 0.3$$
$$\alpha = 0.05$$
$$\theta_s = 0.3, \theta_l = 0.3$$
$$f_p = 0.15, f_n = 0.01$$

A noisy 20x20 pixel image of an edge-detected box was used as an input, representing the retinal P ganglion outputs. Processing was allowed to proceed through 20 iterations. Figure 5 illustrates the input and figure 6 the result, with maximum activations displayed in black.

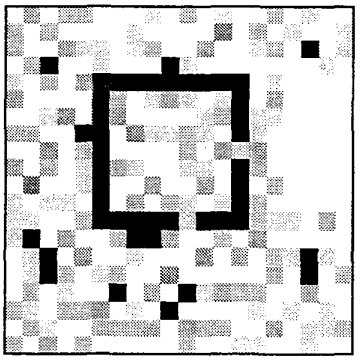

Figure 5. Retinal P ganglion output

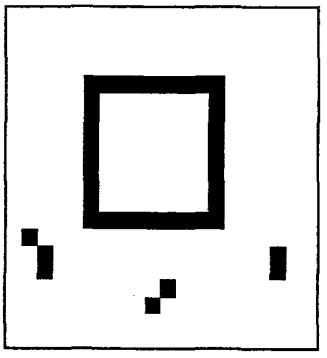

Figure 6. Cortical simple cell output

As can be seen from the figures, the edges in the box have been strengthened and closed, while some of the adjacent noise pixels have been eliminated.

## Conclusions

Beginning with a set of differential shunting equations, a set of steady-state equations has been assembled to model the behavior of the human retina and Lateral Geniculate Nucleus. Examination of the retinal equations has demonstrated that the behavior approximates the Laplacian edge detection method, together with an adaptive low-pass filtering mechanism. While not efficient enough for practical use, the model suggests that the overall form of the human vision model might be adapted to conventional edge detection and adaptive filtering methods to yield similar or improved results, while reducing the computational workload.

The LGN model demonstrated a unique capability for noise reduction that seems to have no counterpart in classical edge detection methods. While reasonably computationally efficient, it has the drawback of a geometrically-increasing workload as the number of orientations is increased. Like the retinal model, it also suggests that with adaptation to the strengths of current image processing hardware, an attractive and powerful enhancement to existing machine vision systems may be possible.

Work with these biologically-inspired models is proceeding to determine if these efficiencies are indeed achievable, as well as to confirm if the techniques will lead to some of the same optical illusions well known to human vision researchers.

## References


[1] Crick, F., *The Astonishing Hypothesis: The Scientific Search for the Soul*, Charles Scribner's Sons, 1994

[2] Enke, D.L., *Biologically Inspired Neural Network Connectionist Models for Use in Artificial Vision Systems*, PhD Dissertation, University of Missouri – Rolla, 1997

[3] Felleman, D.J., and D.C. Van Essen, "Distributed Hierarchical Processing in the Primate Cerebral Cortex", *Cerebral Cortex*, Vol. 1 (1991): 1-47

[4] Gaudiano, P., "A Unified Neural Network Model of Spatiotemporal Processing in X and Y Retinal Ganglion Cells. I: Analytical Results", *Biological Cybernetics*, Vol 67 (1992), 11-34

[5] Gaudiano, P., "Toward a Unified Theory of Spatiotemporal Processing in the Retina", appearing in Carpenter, G., and S. Grossberg (Eds.), *Neural Networks for Vision and Image Processing*, MIT Press, 1992, 195-220

[6] Gaudiano, P., Simulations of X and Y Retinal Ganglion Cell Behavior with a Nonlinear Push-pull model of Spatiotemporal Retinal Processing, *Vision Research*, Vol 34, 1767-1784

[7] Matlin, M.W., and H.J. Foley, *Sensation and Perception*, Allyn and Bacon, 1992